

**We claim:**

1. A system for accessing unstructured objects through by providing structured information, including at least one representative concept for each object, the system comprising:

5 at least one unstructured object contained on a storage medium;  
a procedure that, when executed, accesses the unstructured object and generates structured information including at least one concept for each object; and  
a storage medium containing the structured information.

10 2. The system of claim 1 wherein the system further comprises:

an object concept-based search procedure that, when executed,

accepts search text;

identifies at least one concept in the search text;

accesses the structured information; and

15 identifies objects represented by concepts deduced from the search text.

3. A method for categorizing unstructured objects based on concepts, wherein one object is associated with at least one category, a concept comprises at least one

20 word and a concept grouping comprises related words, the method comprising:

a. building a set of categories, each category comprising at least one word;

b. capturing a set of concepts from the set of categories;

c. expanding the concepts into concept groupings; and

d. indexing the unstructured objects according to the concept groupings.

4. A method according to claim 3, further comprising:

e. selecting at least one key concept for each object based on a statistical procedure; and

5 f. determining at least one category for each object based on a statistical procedure.

5. The method of claim 3, wherein capturing a set of concepts from the set of categories further comprises identifying seed concepts by:

10 extracting words from the set of categories;  
performing a dictionary or thesaurus look-up; and  
generating a list of seed concepts.

6. The method of claim 3, wherein the concept groupings comprise a set of  
15 related words named with a seed concept comprising at least one word, and the set  
of related words include at least one of synonyms, related words, meaning words,  
and user-entered words; and expanding the concepts into concept groupings further  
includes at least one of:

determining synonyms through a thesaurus look-up;  
20 determining related words through a thesaurus look-up;  
determining meaning words through a dictionary look-up; and  
augmenting the set of related words with words entered by a user.

7. The method of claim 3, wherein indexing the unstructured objects according to the concept groupings further includes:

generating a vector to represent an unstructured object where the vector has at least one dimension and each dimension corresponds to one concept grouping;

5 and

determining a weight for each dimension based on a position of words and frequency of occurrence of words in the object.

8. The method of claim 3, wherein selecting at least one key concept for each object based on a statistical procedure includes:

10

determining a probability for each concept grouping represented in an object;

and

selecting key concepts based on the probability determined for the concept grouping.

15

9. The method of claim 7, wherein determining at least one category for each object based on a statistical procedure further includes:

determining a score for each category based on the probabilities of concepts associated with the category; and

20

selecting categories whose score is equal or greater than a threshold value.

10. A method for performing an object concept based search, the method comprising:

determining at least one key concept for each object based on a statistical analysis;

5 accepting one or more user-entered words as a search text;

performing a semantic analysis on the search text to extract at least one extracted concept from search text; and

determining which objects have a key concept matching an extracted concept.

10

11. The method of claim 10, further comprising:

if no matches are determined, performing a keyword-based search and updating concept groupings to include concepts extracted from search text.

15 12. A system for accessing and analyzing unstructured objects, where the system provides structured information through which a user can access the unstructured objects, the structured information including a set of concepts where each concept comprises at least one word, the system comprising:

a first storage medium storing at least one unstructured object;

20 a second storage medium storing a analysis and categorization engine procedure that, when executed, accesses the unstructured objects and generates structured information about the objects;

a third storage medium storing the structured information; and

10087053-030409

a computer processor accessible to the user, having access to the structured information.

13. The system of claim 12, wherein the first, second, and third storage medium  
5 comprise a memory device.

14. The system of claim 13, wherein the memory device comprises a hard disk drive.

10 15. The system of claim 13, wherein the memory device comprises an internet server.

15 16. The system of claim 12, wherein the structured information comprises concepts, each concept having at least one word, extracted from the unstructured objects.

17. The system of claim 12, further comprising:  
a fourth storage medium storing a procedure that, when executed, accepts input from the user, extracts concepts from the input, and performs a search of the  
20 structured information.

18. The system of claim 12, wherein the structured information is stored on the third storage medium in a form of at least one relational database data structure.

19. The system of claim 18, wherein the at least one relational database data structure comprises a relational database table having a global seed concept ID field and a seed concept text field, and a created date field.

5

20. The system of claim 18, wherein the at least one relational database data structure comprises a relational database table having a user ID field, a global seed concept ID field, a related concept ID field, a type of relationship field, and a status field.

10

21. The system of claim 18, wherein the at least one relational database data structure comprises a relational database table having an object ID field, a concept ID field, a cross-reference time stamp field, a cross-reference type field, an index start time field, and a total hits field.

15

22. The system of claim 18, wherein the at least one relational database data structure comprises a relational database table having a user object id field, a key concept id field, a probability field, and a rank field.

20 23. The system of claim 18, wherein the at least one relational database data structure comprises a relational database table having a user ID field, a user object ID field, an object ID field, a user object hierarchy pointer field, and object status field, and an object score field.

24. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer system comprising a analysis and categorization engine to categorize objects based on concepts, wherein one object is associated with at least one category, a concept comprises at least one word and a concept grouping comprises related words, the computer program mechanism comprising:

a program module that directs a analysis and categorization engine to function in a specified manner, the program module including instructions for:

- a. building a set of categories, each category comprising at least one word;
- b. capturing a set of concepts from the set of categories;
- c. expanding the concepts into concept groupings; and
- d. indexing the unstructured objects according to the concept groupings.

25. A computer program product as in claim 24, the program module further including instructions for:

- e. selecting at least one key concept for each object based on a statistical procedure; and
- f. determining at least one category for each object based on a statistical procedure.

26. A relational database data structure comprising:

at least one relational database table having a global seed concept ID field and a seed concept text field, and a created date field.

27. The relational database data structure of claim 26, wherein the at least one relational database data structure comprises a relational database table having a user ID field, a global seed concept ID field, a related concept ID field, a type of relationship field, and a status field.

28. The relational database data structure of claim 26, wherein the at least one relational database data structure comprises a relational database table having an object ID field, a concept ID field, a cross-reference time stamp field, a cross-reference type field, an index start time field, and a total hits field.

29. The relational database data structure of claim 26, wherein the at least one relational database data structure comprises a relational database table having a user object id field, a key concept id field, a probability field, and a rank field.

30. The relational database data structure of claim 26, wherein the at least one relational database data structure comprises a relational database table having a user ID field, a user object ID field, an object ID field, a user object hierarchy pointer field, and object status field, and an object score field.



31. An operating model for concept-based unstructured information gathering and analysis, said operating model comprising:

receiving an unstructured request from a requestor for an information gathering and analysis task;

5 generating a query to gather and analyze unstructured information from at least one identified source;

performing the generated query; and

delivering results of the query to the requestor.

10 32. An operating model as in claim 31, wherein generating a query and performing the generated query include performing a procedure for categorizing unstructured objects based on concepts, wherein one object is associated with at least one category, a concept comprises at least one word, and a concept grouping comprises related words.

15

33. An operating model as in claim 31, wherein the procedure for categorizing unstructured objects based on concepts comprises:

a. building a set of categories, each category comprising at least one word;

b. capturing a set of concepts from the set of categories;

20 c. expanding the concepts into concept groupings; and

d. indexing the unstructured objects according to the concept groupings.

34. An operating model as in claim 33, further comprising:

e. selecting at least one key concept for each object based on a statistical procedure; and

f. determining at least one category for each object based on a statistical procedure.

35. An apparatus for providing structured information about unstructured objects, such that analysis tools for structured information can be utilized to access the unstructured objects, the apparatus comprising at least one storage device storing:

a list of objects;

a list of users;

a list of concepts;

for each concept, a list of related concepts and a classification of a relationship between the concept and each related concept; and

for each user, at least one score for each object accessible to the user based on relevance of a concept to the object.

36. An apparatus as in claim 35, further comprising:

for each user, at least one concept for each object accessible to the user that is representative of that object.